SUBJECTIVE WORKLOAD ASSESSMENT IN A SPATIAL MEMORY TASK
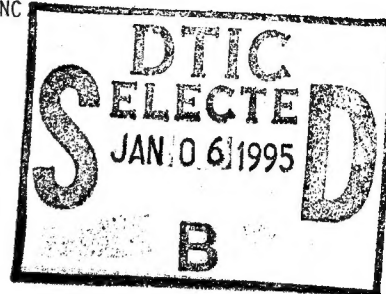
F. Thomas Eggemeier

WRIGHT STATE UNIVERSITY and SYSTEMS RESEARCH LABORATORIES, INC.
Dayton, Ohio

and

Michael A. Stadler

WRIGHT STATE UNIVERSITY
Dayton, Ohio

ABSTRACT

Twelve subjects performed a spatial short-term memory task under several levels of difficulty and rated the workload associated with each using the Subjective Workload Assessment Technique (SWAT). SWAT ratings proved sensitive to two of the three difficulty manipulations in the memory task, and demonstrated greater sensitivity in this respect than either of two primary task measures that were employed. The results extend the applicability of SWAT to the type of spatial memory task used and, therefore, provide further support for the general applicability of SWAT as a workload measurement technique.

INTRODUCTION

A wide variety of assessment techniques have been proposed as measures of operator workload (e.g., Williges and Wierwille, 1979). These techniques vary in their capability to meet the objectives and practical constraints of different applications. Consequently, a comprehensive approach to workload assessment will require the use of several classes of assessment techniques (e.g., Eggemeier, 1984; Shingledecker, 1983).

Subjective measurement techniques are capable of satisfying a number of practical constraints (e.g., ease of implementation, lack of intrusion on operator performance) which must be met in many applications and, therefore, represent a viable candidate for inclusion in a comprehensive workload assessment methodology. Because of their practical utility, subjective techniques have been frequently used to measure operator load in a variety of laboratory and applications environments (Moray, 1982; Williges and Wierwille, 1979).

Despite the widespread use of subjective measures, the rating scale literature has been characterized by individual development of techniques for particular applications. With few exceptions (e.g., Wierwille and Casali, 1983), rating scales have not been evaluated with respect to their sensitivity to workload variations in a variety of tasks. Therefore, there is little evidence in the current literature of a workload rating technique that can be recommended for generalized use as part of an overall workload assessment methodology. In order to provide a workload rating scale with the potential for general applicability, the Subjective Workload Assessment Technique (SWAT) was developed (Reid, Shingledecker, and Eggemeier, 1981a; Reid, Shingledecker, Nygren, and Eggemeier, 1981b).

In SWAT, subjective workload is assumed to be determined by loading on three major dimensions: (1) time, (2) mental effort, and (3) stress. Time load represents the percentage of time that the operator is occupied with information processing requirements, mental effort refers to the degree of concentration and attention required during performance, and stress load represents any additional factors that lead to confusion or anxiety during performance. Each dimension is represented by a 3-point rating scale with verbal descriptors that outline levels on each dimension. SWAT represents an application of conjoint measurement and scaling (e.g., Krantz and Tversky, 1971; Nygren, 1982) which permit ratings on the three dimensions to be combined into one overall interval scale of workload.

Application of SWAT is a two-step process. During an initial scale development phase, subjects rank order the subjective workload associated with all possible combinations of time, effort, and stress load. This ordering information is used to determine the rule for combining the three dimensions into the overall interval scale that results from the procedure. During the second or event-scoring phase, subjects use each of the three individual scales to rate the levels of loading associated with performance of a particular task. The combination of ratings on the three scales is then used to specify a value on the overall interval scale that was derived

1995 0104 084

during scale development. More detailed discussions of the SWAT procedure can be found in Reid et al. (1981a, b).

An important aspect of the SWAT development program has been determining the sensitivity of the technique to workload variations in a number of different types of tasks. Previous investigations have demonstrated that SWAT is sensitive to variations in the difficulty of several laboratory tasks, including critical tracking and simulated aircrew radio communications (Reid et al., 1981a), display monitoring (Notestine, 1983), and verbal short-term memory (Eggemeier, Crabtree, Zingg, Reid, and Shingledecker, 1982). SWAT ratings have also proven sensitive to variations in several forms of task loading in high fidelity flight simulations, such as the number of opponents in aerial combat (Reid, Eggemeier, and Shingledecker, 1984) and the presence or absence of threats to an aircraft (Skelly, Reid, and Wilson, 1983).

The purpose of the present study was to further explore the general applicability of the SWAT technique by examining its sensitivity to a number of difficulty manipulations in a spatial short-term memory task. The requirement to process and remember spatial information is a central component of operator activity in many complex situations (e.g., air traffic control, piloting an aircraft). Also, some current descriptions (e.g., Wickens, 1984) of limits within the human system suggest that spatial information processing draws on capacities/resources that are separate from those used in processing verbal materials. Therefore, it was considered important to determine the sensitivity of SWAT loading to variations in a task which involved spatial processing. The task used in this experiment was a modified version of a pattern recognition task that was originally developed by Chiles, Alluisi, and Adams (1968). The current task required that subjects retain a histogram pattern over a short retention interval and indicate whether a subsequently presented histogram matched the original or "target" pattern. Task difficulty was manipulated by varying the complexity and spatial orientation of the target histogram and by varying the length of the retention interval.

METHOD

Subjects

Subjects (Ss) were five male and seven female introductory psychology students at Wright State University. Ss received extra course credit for their participation.

Apparatus

Memory stimuli were presented on a 12-inch black and white video monitor that was controlled by a Commodore VIC-20 microcomputer. Ss were seated approximately 50 cm in front of the monitor, and responded by pressing one of two buttons on a keypad that was placed a comfortable distance in front of the S's preferred hand.

Procedure

At the beginning of the session, Ss completed the scale development phase of the SWAT procedure. Ss were given a deck of 27 cards, each of which contained statements that represented one of the possible combinations of the time, mental effort, and stress dimensions from the 3-point SWAT scales. Ss sorted the cards so that the 27 combinations were rank-ordered to reflect the degree of subjective workload imposed by each. A Kendall coefficient of concordance ($W = .78$, $p < .01$) demonstrated significant agreement among Ss with respect to rank orderings of the 27 combinations. Therefore, the card sort data from all Ss was combined during the SWAT scaling procedure to form one overall interval scale of workload with a range of 0 to 100. Values from this scale were then used as the measure of subjective load in subsequent analyses.

Each trial in the spatial memory task began with presentation of the target histogram for a 3-second period on the video monitor. The target histogram was then replaced by a three-digit number which remained on the screen for the duration of the retention interval. Ss counted backwards successively by three's from this number in order to prevent any verbal rehearsal of the target pattern information. The counting was done aloud to ensure that Ss performed the subtraction task throughout the retention interval. At the completion of the retention interval, the number was replaced by a second or comparison histogram. Ss indicated whether this histogram was the same or different than the target pattern by pressing the appropriate button on the keypad as quickly as possible. Length and configuration of the bars in the target histogram were chosen at random by the program, as was the choice of whether the second histogram would be the same or different than the first. If it was different, the program again determined the length and configuration of the bars at random. The computer also recorded Ss' responses and reaction times.

Task difficulty was manipulated by varying the complexity of the target histogram (two versus six bars), its spatial orientation (rotations of 0, 90, or 180 degrees from upright), and the length of the retention interval (16 versus 32 seconds). The second or comparison histogram was always presented in the upright position. Retention interval was a between-Ss variable, with six Ss performing the task at each interval. Within

both retention interval groups, each S performed a block of ten trials with each possible combination of the two histogram complexity levels and three orientations for a total of 60 trials. The order of these combinations was counterbalanced across the Ss in each retention interval group. After each block of 10 trials, Ss rated the time, mental effort, and stress load imposed by the combination of complexity, orientation, and retention interval represented in the block.

Prior to actual data collection, Ss received practice on the spatial memory task and on performing SWAT ratings. After a short demonstration of the task, Ss performed blocks of ten practice trials under each of the following conditions: two bars at 0-degree orientation, four bars at each orientation, and six bars at the 180-degree orientation level. The retention interval for all practice blocks was 24 seconds. Ss practiced SWAT ratings during the four-bar conditions, but not in the two- or six-bar conditions since these would be rated during actual testing.

RESULTS

The SWAT rating data were analyzed using a three-factor analysis of variance (ANOVA). Two levels of retention interval (16, 32 seconds), two levels of histogram complexity (two, six bars), and three levels of orientation (0, 90, 180 degrees) were included in the ANOVA. The ANOVA indicated that both the main effect of retention interval $[F(1,10) = 7.57, p < .05]$ and histogram complexity $[F(1,10) = 16.69, p < .01]$ were statistically significant. However, neither the main effect of orientation $[F(2,20) = 1.81, p > .10]$ nor any of the interactions proved to be significant. Figure 1 shows mean SWAT
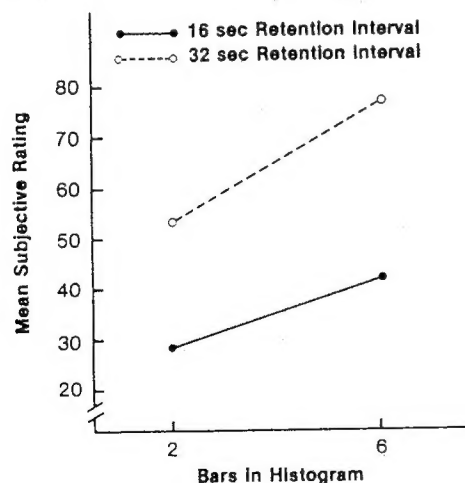
ratings as a function of both the significant retention interval and histogram complexity factors. As is clear from Figure 1, SWAT ratings increased with increases in both levels of complexity and retention interval.

Both memory task errors and reaction times were analyzed with individual 2x2x3 ANOVAs that were comparable to the ANOVA performed on the SWAT rating data. The ANOVA performed on the error data failed to demonstrate any significant main effects or interactions, and so no further analyses were performed on these data. The reaction time ANOVA yielded a significant main effect of histogram complexity $[F(1,10) = 12.69, p < .01]$, but failed to demonstrate significant main effects of retention interval $[F(1,10) = 0.36, p > .25]$ or orientation $[F(2,20) = 3.20, p < .10]$. None of the interactions proved to be statistically reliable. Figure 2 shows mean reaction time as a function of histogram complexity and retention interval and provides a basis to compare the reaction time results with the SWAT ratings in Figure 1. As with the SWAT data, reaction time increased with increases in histogram complexity and, particularly in the six-bar condition, showed a tendency to increase as the retention interval was lengthened.
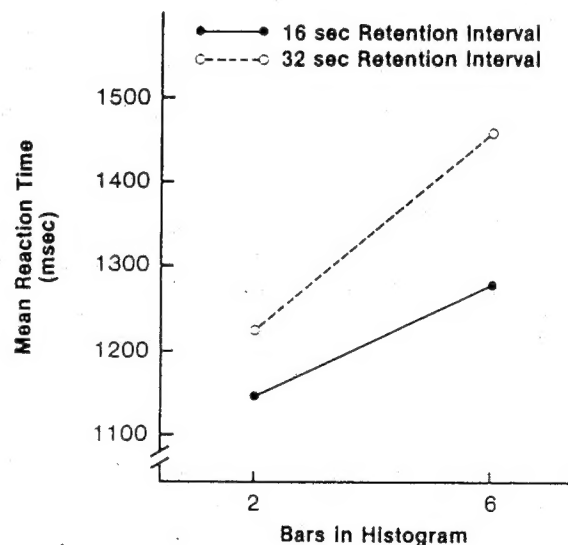
Figure 2. Mean Reaction Time as a Function of Histogram Complexity and Retention Interval

Figure 1. Mean SWAT Ratings as a Function of Histogram Complexity and Retention Interval

DISCUSSION

The results generally provide support for the sensitivity of the SWAT technique to demand manipulations in the type of spatial memory task employed in this experiment. The SWAT procedure proved sensitive to two of the three difficulty manipulations that were used, and demonstrated greater sensitivity in this regard than either of the primary task

measures that were employed. The results, therefore, extend the applicability of SWAT to a type of task that had not been previously researched and, as such, provide further support for the general applicability of the procedure as a workload measurement technique.

Although SWAT failed to discriminate levels within the orientation difficulty manipulation, this was also true of the two primary task measures. Apparently, any additional levels of load that were imposed by this manipulation were lower than those associated with either the complexity or retention interval manipulations. One potential factor which could have minimized the orientation manipulation effect is the fact that while the target histogram appeared in several orientations, the comparison histogram was always presented in the upright position. Therefore, subjects could have adopted the strategy of mentally rotating the target histogram to the upright position before storing the image for later comparison with the second histogram. Some support for the feasibility of this strategy and for its possible effect on the reaction time measure can be derived from previous work on the rotation of mental images (Cooper and Shepard, 1973). Cooper and Shepard (1973) demonstrated that increases in reaction time associated with orientation differences between a target and comparison image could be eliminated by providing advance information to subjects concerning the orientation of the comparison figure. However, the information was effective only if it was provided in time for subjects to mentally rotate the target figure prior to its comparison with the second image. Since subjects in the present study always had advance information about the orientation of the comparison histogram, and since there was sufficient time to rotate the target prior to the comparison operation, it appears that the suggested strategy would have been feasible and would account for the lack of reaction time differences as a function of target orientation. Also, since the rotated images would have been identical to those stored on nonrotation trials, memory storage and retrieval demands assoicated with the orientation variable would have been minimized. Apparently, this minimal additional loading was not sufficient to be reflected in either the error measure or in the SWAT ratings.

The capability of SWAT to discriminate levels of loading imposed by the retention interval variable is noteworthy, since neither the memory error nor reaction time measure demonstrated significant differences as a function of this variable. This type of result is consistent with a major rationale for use of subjective assessment techniques in addition to primary task performance measures, since the latter are capable of discriminating overload from nonoverload situations (e.g., Williges and Wierwille, 1979) but can be insensitive to nonoverload demand manipulations. It appears that the retention interval manipulation used in this experiment was within the nonoverload region of the workload-performance relationship, permitting subjects in the 32-second group to perform at a level comparable to the 16-second group. However, this equivalent performance was apparently achieved at the expense of greater effort or capacity expenditure which was reflected in the more sensitive subjective measure. This high level of SWAT sensitivity relative to primary task performance is consistent with previous results obtained with a verbal short-term memory task in which SWAT discriminated levels of a stimulus presentation rate variable that did not significantly affect memory performance (Eggemeier, Crabtree, and LaPointe, 1983). Therefore, the pattern of results in this and the Eggemeier et al. (1983) experiments suggests that an advantage of SWAT relative to primary task measures is an increased level of sensitivity to some forms of task demand manipulation.

It is important, however, that the noted patterns of sensitivity not be interpreted as indicating that SWAT should replace the use of primary task measures in assessment of operator workload. Primary task measures can provide valuable information about levels of operator performance and permit identification of overloads that result in significant performance degradations. SWAT, on the other hand, can provide valuable information regarding variations in loading and the potential for overloads that is not reflected in primary task performance during nonoverload conditions. The noted differences in sensitivity and in the type of information provided therefore suggest the complementary use of both techniques in a comprehensive approach to operator workload assessment.

ACKOWLEDGEMENT

REFERENCES

Chiles, W. D., Alluisi, E. A., and Adams, O. S. (1968). Work schedules and performance during confinement. Human Factors, 10, 143-196.

Cooper, L. A. and Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), Visual Information Processing. New York: Academic Press.

Eggemeier, F. T. (1984). Workload metrics for system evaluation. Proceedings of the Defense Research Group Panel VIII Workshop "Applications of System Ergonomics to Weapon System Development," Shrivenham, England, C-5-C-20.

Eggemeier, F. T., Crabtree, M. S., and LaPointe, P. A. (1983). The effect of delayed report on subjective ratings of mental workload. Proceedings of the Human Factors Society 27th Annual Meeting, 139-143.

Eggemeier, F. T., Crabtree, M. S., Zingg, J. J., Reid, G. B., and Shingledecker, C. A. (1982). Subjective workload assessment in a memory update task. Proceedings of the Human Factors Society 26th Annual Meeting, 643-647.

Krantz, D. H. and Tversky A. (1971). Conjoint measurement analysis of composition rules in psychology. Psychological Review, 78, 151-169.

Moray, N. (1982). Subjective mental workload. Human Factors, 24, 25-40.

Notestine, J. (1983). Subjective workload assessment in a probability monitoring task and the effect of delayed ratings. Unpublished Masters Thesis, Wright State University, Dayton, Ohio.

Nygren, T. E. (1982). Conjoint measurement and conjoint scaling: A users guide. Wright-Patterson Air Force Base, Ohio: Air Force Aerospace Medical Research Laboratory, Technical Report AFAMRL-TR-82-22.

Reid, G. B., Eggemeier, F. T., and Shingledecker, C. A. (1984). Workload analysis for the AMRAAM operational test and evaluation. Wright-Patterson Air Force Base, Ohio: Air Force Aerospace Medical Research Laboratory (in preparation).

Reid, G. B., Shingledecker, C. A., and Eggemeier, F. T. (1981a). Application of conjoint measurements to workload scale development. Proceedings of the Human Factors Society 25th Annual Meeting, 522-526.

Reid, G. B., Shingledecker, C. A, Nygren, T. E., and Eggemeier, F. T. (1981b). Development of multidimensional subjective measures of workload. Proceedings of the 1981 IEEE International Conference on Cybernetics and Society, 403-406.

Shingledecker, C. A. (1983). Behavioral and subjective workload metrics for operational environments. Proceedings of the AGARD (AMP) Symposium and Sustained Intensive Air Operations: Physiological and Performance Aspects, AGARD-CP-338, 6-1-6-10.

Skelly, J. J., Reid, G. B., and Wilson, G. R. (1983). B-52 full mission simulation: Subjective and physiological workload applications. Paper presented at the Second Aerospace Behavioral Engineering Technology Conference, Long Beach, California.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman and R. Davies (Eds.), Varieties of Attention. New York: Academic Press.

Wierwille, W. W. and Casali, J. G. (1983). A validated rating scale for global mental workload measurement applications. Proceedings of the Human Factors Society 27th Annual Meeting, 129-133.

Williges, R. C. and Wierwille, W. W. (1979). Behavioral measures of aircew mental workload. Human Factors, 21, 549-574.